DOCUMENT RESUME

ED 179 295 PS 011 029

AUTHOR Horn, John L.
TITLE Statistical Weaknesses of a Unitary Construct of Intelligence.
PUB DATE Sep 79
NOTE 15p.: Paper presented at the Annual Meeting of the American Psychological Association (87th, New York, NY, September 1-5, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Factor Analysis; Intellectual Development; *Intelligence; *Intelligence Factors; *Intelligence Quotient; *Intelligence Tests; Research Needs; Research Problems; Testing Problems; Theories
IDENTIFIERS General Intelligence; *Wechsler Adult Intelligence Scale

ABSTRACT

A brief discussion of theories of general intelligence precedes a report of relevant empirical data. Results from the factor analysis of more than 20 sets of Wechsler Adult Intelligence Scale (WAIS) data indicate that the WAIS clearly is not a one-factor scale. It does not measure a single, general intelligence. Roughly 17 percent of the reliable variance of the WAIS remains unaccounted for by a three-factor model. There must be at least three, and probably at least four, major components responsible for the individual differences that are represented by the total IQ obtained with the WAIS. Additional analyses on other data suggest that as many as eight major factors are needed to account for the individual differences measured by intelligence tests. General intelligence and the IQ cannot suffice as a basis for a science of human abilities because they have not been defined in terms that are objective enough to ensure that the same phenomena can be observed under different conditions of observation. There is, however, a wide range of phenomena that to some extent can be described by use of a concept of general intelligence and a measure of IQ; IQ measures of general intelligence do provide a crude basis for making diagnoses and decisions. Future studies of intelligence, though, should sample a wide domain of behaviors across the life span and across cultures. Exclusive reliance on a single measure should be avoided.
(Author/RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED179295

STATISTICAL WEAKNESSES OF A UNITARY CONSTRUCT OF INTELLIGENCE

APA Symposium Presentation, September 5, 1979

John L. Horn

University of Denver

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Introduction: The Power of the IQ

There are few things psychologists enjoy more than a feast of criticism of the IQ. Such feasts have been conducted frequently over the last sixty years. It's almost a ritual of the profession to solemnly convene at some annual meeting or another, or a specially arranged conference, to roundly condemn the poor old IQ.

Yet, curiously, the concept survives. After the dust has cleared from each ferocious encounter of criticism, the IQ emerges, Phoenix-like, tattered a bit and showing the signs of battle, but standing upright ready to continue doing service until the next occasion when psychologists and educators joyously convene to again righteously batter the concept and the measure. There is virtually no side of the IQ that has not been severely and cogently criticized. Yet it remains with us: why?

Well, one part of the answer to this question is that the IQ represents a theory about an important feature of human functioning. A theory, no matter how bad, is not discarded, and scientists know it should not be discarded, until there is a better theory to put in its place; it seems that over the last 60 years there has been no better theory to put in place of the IQ. Perhaps even today there is no theory that can truly supplant the theory of the IQ.

At this meeting such statements may constitute dangerous heresy. Should I beware lest you stone me from the podium? How can one dare to say that the lowly IQ has properties as laudable as those of a genuine theory? Where are these properties? What important phenomena are in the least explained by ideas about the IQ? How do these ideas aid our scientific understanding of human behavior?

PS 01 1029

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
John L. Horn
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

To answer these questions we should first recognize that the IQ represents concepts of general intelligence and then note that one reason the IQ has remained with us is because theories about general intelligence have provided a reasonable basis for understanding human abilities. When we look to see what it is that these theories explain, we can see the research and theoretical discussions of such people as Humphreys (1976; 1979) and Jensen (1973; 1979) and Matarazzo (1972) as well as the writings of test constructors, such as Wechsler (1974).

We need to consider the strengths of theories about general intelligence in order to lay a foundation for understanding the hardiness of the IQ and in order to understand how we might replace the theories of general intelligence and the IQ with a better set of concepts (and tests).

Strengths of Theories of General Intelligence

What, then, are the important phenomena that are explained in part by theories about general intelligence?

First, there is the persistent finding of positive manifold among the intercorrelations for almost all of a vast number of ability performances. It seems that almost all performances that can be accepted as indicative of quality in thinking are supported by a "something" that is possessed more by some people than by others. This "something" is also measured with IQ tests; it is measured imperfectly, to be sure, but nevertheless with some accuracy. This "something", which we refer to as general intelligence or IQ, represents an important, pervasive feature of human abilities.

A second set of evidence that is important for theories of general intelligence is a set of replicated and reasonably consistent results suggesting that a general factor is determined, in part, by genetic factors in which there are, collectively, individual differences. There is also evidence indicating a developmental simplex

for the general factor. The reference here is to Anderson's early work and the more recent work of Humphreys (1960; 1979) and his coworkers.

A variety of other evidence is also consistent with a theory of general intelligence. For example, most of the data on social class and gender differences pertains to the IQ rather than to any set of constructs that might reasonably be thought to be a substitute for the IQ. Humphreys (1979) has shown that individual differences in schools are most fully accounted for by a general factor that is similar to an omnibus measure of IQ. Results pertaining to laboratory tasks designed to indicate cognitive functioning and information processing, as in the work of Carroll (1976; 1978), Hunt (1978), Jensen (1979), and Sternberg (1977), have often been discussed as if they related primarily to a general factor of intelligence. In the practical business of selection and placement it has been recognized that often the safest bet to yield a noteworthy and dependable validity coefficient is an omnibus measure of IQ, as McNamar (1964) cogently argued several years ago. Often in clinical diagnosis it is the general factor that seems to be most useful in understanding how intellectual functioning relates to the problems a patient is experiencing. When there is damage to the brain, no matter where this occurs, it seems to have at least some effect on the totality of intellectual functioning; such effects are detected, however imperfectly, with IQ tests.

In general, then, there is a wide range of phenomena that we can describe, to some extent, by use of a concept of general intelligence and a measure of the IQ. We can criticize the fact that this description is far from precise, far from adequate for many purposes, and far from what we would like it to be, but we should not gainsay the fact that the description is better than nothing and often better than what we get with other concepts and measures. We need to keep this perspective as we look to see where we fail when we persist in using a concept of general intelligence and a measure of the IQ.

Weaknesses of Theories of General Intelligence

One major weakness of general intelligence theories and IQ measures is their chameleon quality: they change color, so to speak, with changes in the situations in which they are used. A test used at different age levels measures somewhat different factors. The Stanford-Binet, for example, measures different abilities when used with five-year old children than when used with young adults. Even when used at merely different ability levels, IQ tests measure somewhat different patterns of abilities and capacities. The Stanford-Binet measures different abilities when used with adult retardates than when used with normal adults. Different tests designed to measure the same general factor have somewhat different factor compositions. The Stanford-Binet measures somewhat different abilities than the Wechsler scales.

These problems are not solved by letting the general factor among a collection of ability performances define the concept of general intelligence. Indeed, the problems are illuminated by this procedure! For unless there is a clearly formulated prescription for sampling ability performances, the general factor can fluctuate a great deal with each different sampling of ability tasks and subjects.

A difficulty with theories of general intelligence is that they do not provide a clearly formulated prescription for sampling the capacities of intelligence. What are the necessary and sufficient indicants of the IQ? What subtests must be included to measure it properly and what tests can be omitted either because they are redundant or because they don't measure an essential aspect of the IQ? How much weight should be allotted to one subtest in comparison to another in order to achieve the correct balance in a measure of the general factor?

We don't have good answers to these questions and because we don't have good answers, we really don't know how to measure the IQ. For this reason each investigator, and each merchant of psychological testing wares, can have a different measure of IQ, and it is virtually impossible to argue, on the basis

of theoretical principles, that one of these measures is any better or any worse than another. Moreover, we have no concepts, analogous to the laws of thermodynamics in physics, that enable us to maintain that the same "something", regulated by the same laws, is manifested in somewhat different ways when measured in the various different ways of measuring the IQ.

This difficulty with the concept and measure of general intelligence is a major problem with, for example, Jensen's (1973) efforts to demonstrate the heritability of the IQ. In his book on Educability and Group Differences, Jensen collates results from a large number of studies. It is a worthy effort. But after it is done we are not sure whether the evidence in one section of the book, based on one test, is even relevant to, much less consistent with, the evidence of another section of the book, based on another test. The complexion of IQ is different in each section and we cannot be sure which hue is being reflected in the mirror of the evidence that Jensen cites.

Jensen's efforts to use the concept of general intelligence are better than many others. His failure can thus be viewed as generic. General intelligence and the IQ cannot suffice as a basis for a science of human abilities because they have not been defined in terms that are objective enough to ensure that the same phenomena can be observed under different conditions of observation.

Some Relevant Empirical Data

Results from some of our recent work with the Wechsler Adult Intelligence Scale (WAIS) provides a concrete basis for considering the multiple-component nature of the IQ. In Table 1, for example, there is a summary of some analyses indicating that the WAIS is comprised of several independent factors. Using maximum likelihood estimation and statistical evaluation procedures developed by Jöreskog and Sörbom (1978), Jack McArdle and I have done a number of model fit analyses for more than 20 sets of WAIS data (each based on sample sizes of 150 or

Table 1

Structure of the WAIS Examined By

Maximum Likelihood Fits For Different Hypotheses

(After McArdle and Horn, 1979)

		General Intelligence		Verbal and Performance			Gf/Gc Theory				
		G	$r^2_{xt} - h^2$	Verbal	Perform	$r^2_{xt} - h^2$	Gc	Gf	SAR	$r^2_{xt} - h^2$	r^2_{xt}
Information	IN	.85	.19	.88		.14	.92			.07	.91
Comprehension	CO	.77	.19	.78		.17	.83			.10	.78
Arithmetic	AR	.69	.33	.69		.33	.23		.76	.12	.81
Similarities	SI	.77	.26	.76		.27	.68	.15		.22	.85
Memory Span	MS	.57	.34	.55		.36			.69	.19	.66
Vocabulary	VO	.85	.23	.89		.16	.92			.10	.95
Digit Symbol	DS	.62	.54		.67	.47		.26	.49	.41	.92
Picture Completion	PC	.74	.30		.80	.21	.34	.51		.23	.85
Block Design	BD	.66	.39		.76	.25		.80		.19	.83
Picture Arrangement	PA	.67	.21		.74	.11	.39	.50		.12	.66
Object Assembly	OA	.56	.37		.66	.25		.77		.10	.69
Average Reliable Uniqueness			.30			.25				.17	
Chi-Square for Fit			691			295				73	
Degrees of Freedom			44			43				36	
Chi-Square Difference						396				222	
Degrees of Freedom Difference						1				7	

Note: $r^2_{xt} - h^2$ represents the difference between the reliability, representing a variable's squared correlation with its own true score component, and the communality, representing a variable's squared multiple correlation with the common factor components of the other variables of the set. Thus $r^2_{xt} - h^2$ is the true score variance not accounted for by common factors.

The chi-square difference with one degree of freedom represents a nested-set. This difference indicates that the two-factor solution provides a better fit than the one-factor solution. The three-factor solution is similarly better than the two-factor solution but because the latter is not a one df nested set of the former, there may be other three-factor solutions that provide as much chi-square difference between this or other two-factor solutions as the present solutions, and with the same degree of freedom.

larger). The results shown in Table 1 are typical of what we have found in this (as we are regarding it) pilot research.

Insert Table 1 About Here

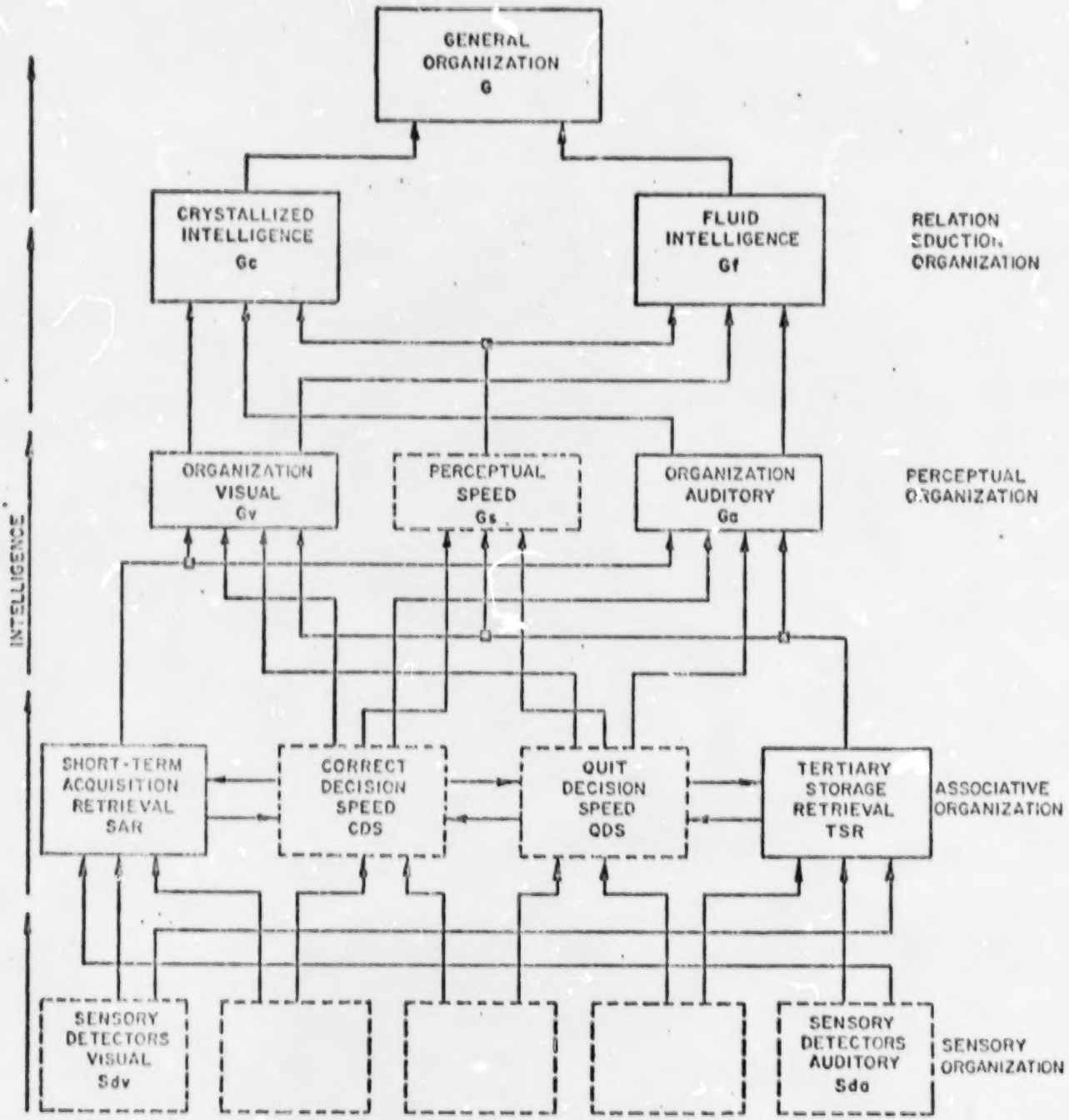
We find that in no case can we even come close to fitting a one-factor model or any two-factor model yet conceived by us or recommended by others, including the verbal/performance two-factor model that is part of the standard scoring of the Wechsler tests. In other words, the WAIS clearly is not a one-factor scale; it does not measure a single general intelligence. Indeed, more than two factors must be invoked to account for the reliable variance of the test scores.

In the work we have completed so far in this line of research we have fitted, or have come very close to fitting, a three-factor model that derives from a general theory of fluid and crystallized intelligence. A typical example of this fit is shown in the right section of Table 1. Here it can be seen that even this fit leaves statistically significant and reliably measured variance yet to be specified by a model: the chi-square, indicating departure from fit, is still roughly two times the degrees of freedom, and $r^2_{xt} - h^2$, indicating the reliable variance that is not accounted for by communality variance, is still .17 on the average. That is, roughly 17 percent of the reliable variance of the test is not accounted for by the three-factor model. It is reasonable to suppose that a four-factor model will account for the residual reliable variance, but four-factor models that we have tried to fit have not been stable enough from one sample of subjects to another to warrant a conclusion that we know how to specify the reliable variance that is left over after we fit a three-factor model.

From analyses such as those summarized in Table 1 we conclude that there must be at least three, and probably at least four, major components, or processes, or forms of intelligence, responsible for the individual differences that are represented by the total IQ obtained with the WAIS. The IQ obtained with this test is thus not unitary at all, but divisable into at least three or four reliably

FIGURE 1

SCHEMATIC REPRESENTATION OF INTELLECTUAL FUNCTIONS



N.B. INTERPRET ARROWS DEVELOPMENTALLY RATHER THAN AS AT-TIME-OF-MEASUREMENT DETERMINATION.

independent components.

These kinds of analyses only scratch the surface of the multiple-factor complexity of the measurements that can be obtained with various kinds of tests and tasks that are indicative of the phenomena of human intelligence. The behavior that Renée Fuller has described at this symposium, as measured by the Stick and Ball task, is also indicative of an important aspect of intelligence, and this factor is almost certainly not well represented in our usual measures of IQ. In some of our work we have analyzed as many as 40 distinct tasks that one researcher or another has regarded as indicative of intelligence. From analyses that are similar to those summarized in Table 1, we have concluded that no fewer than 8 major factors must be invoked to account for, and describe, the individual differences that are measured with tasks that reasonably can be regarded as evoking expressions of human intelligence. A summary of some of our results is indicated by Figure 1.

Insert Figure 1 About Here

Why An IQ Measure Often Suffices

Thus a major weakness of theories and practices that assume a unitary concept of intelligence is that the phenomena to which these theories and practices refer simply are not unitary. To recognize this weakness is not to deny that one can often make do, as it were, with measures that treat this non-unitary phenomena of intelligence as if they were unitary. It's as if one were to weigh collections of pennies, nickels, quarters and half-dollars on an apothecary's scale, and regard the number of ounces or pounds as a measure of amount of money. Financial transactions conducted with such a measure usually would not match the transaction conducted using the values of the coins, but there would be a rough correspondence between the two kinds of transactions and sometimes the exchange based on the weight measure would be precisely the same as the exchange based on use of the values of the coins. The weight measure would thus provide a crude basis for conducting transactions that could be conducted more precisely using knowledge

of the values of the coins. Similarly, IQ measures of general intelligence provide a crude basis for making diagnoses and decisions that could be made more precisely if we knew how to use our knowledge that several different components contribute to the IQ measure. Unfortunately, however, our science is still at a stage of trying to identify and distinguish the things that are analogous to the penny, the nickel, the quarter, and the half-dollar of our example. We know that there are at least eight different "coins" in our conglomerate IQ measure but we don't know precisely the values of the separate "coins."

Why A Statistical Weakness And How to Build Strength

The fact that the concept and the measures of general intelligence are not unitary is a fundamental weakness. Other weaknesses pale into insignificance by comparison. The weakness is essentially one of theory. Why, then, you might ask, should I refer to the weakness as statistical?

My answer to this question is that I believe substantive theory about human intelligence must be based on sampling concepts and procedures that are analogous, at least, to the sampling concepts and procedures of statistical theories. That is, I think we must have research in this topic area that derives from the kind of metatheory that Cattell has been most prominent in advancing, under such headings as domain sampling of the personality sphere (Cattell, 1950) or that Brunswick (1956) has discussed under very similar headings.

This kind of theory requires that one consider several kinds of tests and tasks that can be indicative of intelligence. Such theory is quite different from that which leads to the common practice of considering only one kind of test.

I don't think we can make much progress in developing a truly scientific theory about intelligence with research based on particular tests. This is the major error of our previous research on intelligence. Psychologists have worked competitively for over sixty years to invent new intelligence tests. This program of research has yielded a huge number of ingenious tests (and many self-serving,

exaggerated claims in the literature of intelligence testing) but it has contributed very little to the architectonic development of a scientific theory of intelligence. I fear we could go on for another sixty years inventing all manner of yet other interesting tests and still not have a theory of intelligence with which the profession could be reasonably comfortable (and could present to the general public without the effusive apologies that now accompany most psychologists' communications about intelligence).

There are, of course, very real limitations in theories that have come from factor analytic research on human intelligence. Nevertheless, this research has contributed to such progress as we have made in bringing theoretical order out of the chaos of controversy about the nature of intelligence. A basic feature of this research is the concept of sampling from a domain of behaviors that are, by one criterion or another, indicative of different aspects of intelligence.

Research that aims to describe the components of intelligence, as represented by Sternberg's (1977) popular work, is quite consistent in principle with the domain sampling approach. To identify essential components one must properly sample them. There is a danger, however, that componential analyses will focus on particular tasks--as the analogies test, for example--and thus will fall into the trap of narrowness that characterizes other research on developing and describing a single measure.

The problem in this respect is perhaps seen more clearly in research on personality in general than in research on intelligence. Our profession has made massive, but largely wasteful, institutions out of study of such tests as the MMPI and Rorschach, for example. The research literature contains more information about these tests than anyone should ever want to use, but this information has contributed--and can contribute--only very little to general personality theory. The same thing can happen with componential analyses unless the results are extended to tasks other than merely analogies tests and woven into the fabric of theory

that emerges from study of varied samples of different kinds of ability tests.

Problems With Domain Sampling Theories

The problems with the approach I have advocated should not be underestimated. A notable complexity is that we must somehow conceive of ways of circumscribing this domain. What lies within and what is outside the borders of the domain of truly intellectual abilities? We don't know how to answer this question. We know that we must consider the total life span, and thus the capacities of the youngest infant and the oldest adult. Also, we know that if results we obtain at one time are to be at all applicable to future times, then probably we must sample abilities as they are expressed in all cultures of the world, for this variety provides the only practical base we have for estimating the variety that can emerge as historical time unfolds. A major reason for doing cross-cultural research is to provide a basis for theory that has a chance of being valid and applicable in the future for our rapidly changing cultures.

Summary

The concept of general intelligence is not moribund, but it is weak relative to the kind of theory we should have. It is weak primarily because the phenomena is comprised of several distinct capacities that combine in largely unknown ways in expressions we recognize as human intelligence. Future research must be directed at providing a comprehensive sampling of the abilities of intelligence. This research can provide a basis for identifying the components of intelligence. Such research should be coupled with analyses to determine the modes of combination of components in intelligence. We should not continue to focus as much research as we have focused on the study of single tests and procedures.

References

Brunswick, E. Perception and the representative design of psychological experiments, Berkeley, California, University of California Press, 1956.

Carroll, J. B. Psychometric tests as cognitive tasks: a new "structure of intellect." In the Nature of Intelligence, L. B. Resnick (Ed.): Hillsdale, N. J.: Erlbaum, 1976, 27-56.

Carroll, J. B. How shall we study individual differences in cognitive abilities--methodological and theoretical perspectives. Intelligence, 1978, 2, 87-115.

Cattell, R. B. Personality. New York: McGraw-Hill, 1950.

Humphreys, L. G. Investigations of the simplex. Psychometrika, 1960, 25, 313-323.

Humphreys, L. G. A factor model for research on intelligence and problem solving. Chapter 19 in The Nature of Intelligence. L. B. Resnick (Ed.) Hillsdale, N.J.: Erlbaum, 1976.

Humphreys, L. G. The construct of general intelligence. Intelligence, 1979, 3, 105-120.

Humphreys, L. G., Parsons, C. K. and Park, R. K. Dimensions involved in differences among means of cognitive measures. Journal of Educational Measurement, 1979, 16, 63-76.

Hunt, E. B. Mechanics of verbal ability. Psychological Reviews, 1978, 85, 109-130.

Jensen, A. R. Educability And Group Differences. New York: Harper and Row, 1973.

Jensen, A. R. g: outmoded theory or unconquered frontier? Creative Science and Technology, 1979, 2, 16-29.

Jöreskog, K. G. and Sörbom, D. LISREL-IV: Analysis of linearstructural relationships by the method of maximum likelihood. Chicago: National Educational Resources, 1978.

Matarazzo, J. D. Wechsler's Measurement And Appraisal of Adult Intelligence. Baltimore: Williams and Williams, 1972.

McNamar, Q. Lost: our intelligence? Why? American Psychologists, 1964, 19, 871-882.

Sternberg, R. J. Intelligence, Information Processing And Analogical Reasoning: The Componential Analysis Of Human Abilities. Hillsdale, New Jersey: Erlbaum, 1977.

Wechsler, D. Wechsler Intelligence Scale For Children - Revised. New York: The Psychological Corporation, 1974.